



(11) Publication number:

**0 079 739**  
**A2**

(12)

# **EUROPEAN PATENT APPLICATION**

(21) Application number: **82305926.6**

(22) Date of filing: **08.11.82**

(51) Int. Cl.<sup>3</sup>: **C 12 N 15/00**  
**C 12 N 1/00, C 12 P 21/02**  
**C 07 H 21/04, C 07 C 103/52**  
**//C12R1/19, C12R1/865**

(30) Priority: **12.11.81 US 320632**

(43) Date of publication of application:  
**25.05.83 Bulletin 83/21**

(84) Designated Contracting States:  
**BE CH DE FR GB IT LI NL SE**

(71) Applicant: **THE UPJOHN COMPANY**  
**301 Henrietta Street**  
**Kalamazoo, Michigan 49001(US)**

(72) Inventor: **Dugalczyk, Achilles**  
**c/o The Upjohn Company 301 Henrietta Street**  
**Kalamazoo Michigan 49001(US)**

(74) Representative: **Perry, Robert Edward et al,**  
**GILL JENNINGS & EVERY 53-64 Chancery Lane**  
**London WC2A 1HN(GB)**

(54) **Albumin-based nucleotides, their replication and use, and plasmids for use therein.**

(57) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

**EP 0 079 739 A2**

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION  
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly  $\alpha$ -fetoprotein, but the synthesis decreases drastically after birth. Recently, 10 Law et al determined the complete sequence of mouse  $\alpha$ -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the  $\alpha$ -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T<sup>C</sup> T C T T C T G T.....albumin mRNA  
 (3')...G A G G A A G G C G U C C m<sub>2</sub><sup>6</sup>A m<sub>2</sub><sup>6</sup>A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre-peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro-peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

30

35

TABLE 1

[illegible]

5'	10	15	20	25	30	35
231 val GTT	245 leu TCC	246 cys AAG	247 thr TTC	248 his ACC	249 leu ACA	250 asp GAT
261 ala GCC	262 tyr TAT	263 asn ATC	264 gln GAT	265 asp TGC	266 thr GAT	267 val GAT
291 ala GCC	292 val GTC	293 gln GAT	294 met GAT	295 phe ATG	296 thr GAT	297 leu GAT
321 glu GAG	322 ala GCA	323 lys AAG	324 asp GAT	325 val GTC	326 thr GAT	327 leu GAT
351 lys AAG	352 thr ACA	353 tyr TAT	354 gln GAT	355 cys TGC	356 lys AAG	357 thr GAT
381 val GTC	382 glu GCA	383 pro GAT	384 asn GAT	385 leu TTC	386 thr GAT	387 leu GAT
411 tyr TAC	412 lys ACC	413 val AAA	414 gln GAT	415 thr GAT	416 pro GAT	417 leu GAT
441 pro CCT	442 ala GCA	443 lys GCA	444 arg ATG	445 met GAT	446 pro GAT	447 cys GAT
471 asp CAC	472 arg ACA	473 val ACC	474 thr GAT	475 cys TGC	476 thr GAT	477 leu GAT
501 glu GAG	502 phe TAT	503 asn TAT	504 ala GAT	505 his TTC	506 thr GAT	507 leu GAT

35 531 540 550 558 559 560  
 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala phe val glu lys cys lys  
 GAG CTC GTC AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CTC AAA GCT GTT ATG CAT GAT TTC GCT GCT TTT GTA GAG AAG TCC TGT AAC (1790)  
 561 567 570 580  
 ala asp asp lys glu thr cys phe ala glu glu gln lys leu val ala ala ser gln ala leu gly leu ter  
 CCT GAC GAT AAG GAG ACC TCC TTT CCC GAG GAG GGT AAA AAA CTT GCT GCA AGT CAA GCT GCC TTA TTA TAA CATCACATTAAAG (1883)  
 ter ter  
 CATCTCAGCCTACCATGAGATACAGAGAAAAATCAAGATCAAAAGCTTATTCACTGTTTTCTTTTCTGCTGTAAGCCACACCCCTCTAAACACATAAATTCTTTAA (2002)  
 TCATTTCCTTCCTGCTGCTTCAATTAAATAAAAAATGGAAGAACTAA..... 20 .....AA (2078)

Following are examples which illustrate procedures, ~~including the best mode~~, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1      Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and  
10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685].  
15

Example 2      Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczky, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Bolivar, F.,  
20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Royer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczky, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., Ibid.]. The albumin clones were selected using the colony  
25 hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [<sup>32</sup>P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HB101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to  
35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are



prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3      Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

20 Example 4      DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ[<sup>32</sup>P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

30 Example 5      Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5

HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public, ~~upon the grant of a patent. It should be understood that the availability~~  
10 of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL  
15 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEpl6 is a well known and widely available yeast episomal plasmid.  
20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEpl6) - NRRL B-12093.

Example 6      Assembly of the Serum Albumin Gene

Assembling the pieces together is a straightforward task of re-  
25 striction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the  
30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

35

(a) Escherichia coli

(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRI DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco RI cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.R. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

#### 15 Example 7      Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

#### 30 Example 8      Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies  
5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T.  
10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number  
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number  
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

35  
 30  
 25  
 20  
 15  
 10  
 5  
 -10  
 -1 -6 p f o -1 1  
 ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu aly glu asn phe lys 20  
 TCG CCT TAT TCC ACG CGT CTC TTT CGT CCA CAT CCA CAC AAG AGT CAG GTT CCT CAT CCG TTT AAA GAT TTC GCA CAA GAA AAT TTC AAA (170)  
 21  
 ala leu val leu ile ala phe ala gln tyr leu gln cys pro phe glu asp his val lys leu val asn glu val thr alu phe ala 50  
 GCC TTC CTC TTG ATT GCC TTT CCT CAG TAT CTT CAG CAG TGT CCA TTT CAA CAT CAT GTA AAA TTA GTC AAT CAA GTA ACT CAA TTT GCA (260)  
 34  
 30  
 60 62  
 51 53  
 lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu 80  
 AAA ACA TGT GTT GCT CAT CAG TCA GCT CAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TCC ACA CTT GCA ACT CTT (350)  
 81  
 arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro 110  
 CGT CAA ACC TAT CGT CAA ATG CCT CAC TCC TGT CCA AAA CAA CAA CCT CGG ACA AAT CAA TCC TTC CAA CAC CAA GAT CAC AAC CCA (480)  
 120 124  
 111  
 asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu alu thr phe leu lys lys tyr leu try 140  
 AAC CTC CCC CGA TTC CTC ACA CCA CAG GTT CAT GTG ATG TCC ACT GCT TTT CAT GAC AAT CAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (330)  
 150  
 141  
 glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr alu cys cys gln 168 169 170  
 GAA ATT CCC ACA ACA CAT CCT TAC TTT TAT CCC CCG GAA CTC CTT CAG CCA AAG CTC CAT CAA CTT CGG CAT CAA CGG AAG GCT TCG TCT CCC AAA CAG ACA CTC AAG TGT (420)  
 177 180  
 171  
 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp alu gly lys ala ser ser ala lys gln arg leu lys cys 200  
 GCT CCT CAT AAA GCT CCC TCC CTC TTG CCA AAG CTC CCA AAG CTC CAT CAA CTT CGG CAT CAA CGG AAG GCT TCG TCT CCC AAA CAG ACA CTC AAG TGT (710)  
 210  
 201  
 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala alu phe ala alu 230  
 CCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC CCA TCG CCA GTA CCT CCC CTG ACC CAG ACA TTT CCC AAA CCT CAG TTT GCA CAA (300)

231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA CTC CAC ACC GAA TCC TCC CAT GCA GAT CTG CTT GAA TGT GCT GAT CAC AGC GCG GAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile  
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTC AAG CAA TCC TGT CAA AAA CCT CTG TTG CAA AAA TCC CAC TCC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
 CCC CAA GTG CAA AAT GAT CAG ATG CCT GCT GCT TCG CCT TCA TTA CCT GCT CAT TTT GTT GAA ACT AAG CAT GTT TCC AAA AAC TAT GCT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
 CAG GCA AAG GAT GTC TTC TTG CCG ATG TTT TTG TAT GAA TAT GCA ACA AGG CAT CCT GAT TAC TCT CTC CTG CTG CTG ACA CTT GCC (1160)

351 lys thr tyr glu thr leu glu lys cys oys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
 AAG ACA TAT GAA ACC ACT CTA CAG AAG TCC TGT CCC GCT GCA GAT CCT CAT GAA TCC TAT CCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu qln leu gly glu tyr lys phe qln asn ala leu leu val arg  
 GTG CAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CAG CTT TTT CAG CAG CTT GCA CAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CTT (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA AGA AAC CTA GCA AAA GTG CCC AGC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTC AAC CAG TTA TGT GTC CAT CAG AAA ACG CCA GTA AGT (1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
 CAC ACA GTC ACC AAA TCC TGC ACA GAA TCC TTG CTC AAC AGC CGA CCA TCC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val  
 CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

5  
 10  
 15  
 20  
 25  
 30  
 35

531  
 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala phe val glu lys cys lys  
 CAG CTC GTG AAA CAC CAC CAC AGC CCC AGC GCA ACA AAA GAG CAA CTC AAA GCT GTT ATG CAT CAT TTC GCT TTT GTA GAG AAG TGC TGC AAG (1790)

540  
 550  
 558 559 560  
 561  
 ala asp asp lys glu thr cys phe ala glu glu gln lys lys leu val ala ala ser gln ala ala leu gln leu ter  
 CCT CAC CAT AAG GAG ACC TGC TTT GCC CAG CAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCATTTTAAAG (1883)

ter ter  
 CATCTACGCTACCATGACATTAACAGAAAGAAATGAACATCAAAAGCTTATTTCATCTGTTTTTCTTTTCGTTGCTGTAAGCCACACCCCTGCTCTAAAAACATAAATTTCTTTAA (2002)

TCATTTTGCCTCTTTTCTCTGCTGCTTCAATTAAATAAATAATGGAAGCATCTAA..... 20 .....AA (2078)



6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35
1
asp ala his lys ser glu val ala arg phe lys asp leu gly glu glu asn phe lys
CAT GCA CAC AAG AGT CAG GTT CCT CAT CCG TTT AAA CAT TTG CCA GAA GAA AAT TTC AAA (170)
20
21
ala leu val leu ile ala phe ala gln tyr leu gln gln oys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
CCC TTG GTG TTG ATT GCC TTT CCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)
30
31
51
53
lys thr oys val ala asp glu ser ala gln asp oys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu
AAA ACA TGT GTT CCT GAT CAG TCA CCT GAA AAT TGT CAG CAC AAA TCA CTT CAT ACC CTT TTT CCA GAC AAA TTA TGC ACA GTT GCA ACT CTT (350)
75
80
81
arg glu thr tyr gly glu met ala asp oys oys ala lys gln glu pro gly arg asn glu oys phe leu gln his lys asp asp asn pro
CGT GAA ACC TAT GGT GAA ATG CCT GAT TGC TGT CCA AAA CAA CAA CCT CCG ACA AAT GAA TGC TTC TTG CAA CAC CAC AAA CAT GAC AAC CCA (440)
100
101
111
asn leu pro arg leu val arg pro glu val asp val met oys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
AAC CTC CCC CGA TTG GTG ACA CCA CAC GTT CAT GTG ATG TGC ACT CCT TTT CAT GAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (530)
120
124
130
141
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu oys oys gln
CAA ATT CCC ACA ACA CAT CCT TAC TTT TAT CCG CCG CAA CTC CTT TTC TTT CCT AAA AGC TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA (620)
150
160
168
169
171
ala ala asp lys ala ala oys leu leu pro lys leu asp glu leu arg asp glu gln gln lys ala ser ser ala lys gln arg leu lys oys
CCT CCT GAT AAA GCT GCC TGC CTG TTG CCA AAG CTC CAT GAA CTT CCG CAT CAA CCC AAG CCT TCG TCT CCC AAA CAG ACA CTC AAC TGT (710)
177
180
190
201
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala ala phe ala glu
CCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC CCA TGG GCA GTA CCT CCC CTC ACC CAG ACA TTT CCC AAA GCT CAG TTT GCA GAA (800)
210
220
230

```

231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260  
CTT TCC AAG TTA CTG ACA CAT CTT ACC AAA CTC CAC ACG CAA TCC TCC CAT GCA CAT CTG CTT CAA TGT CCT CAT CAC ACG CCG CAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile 289 290  
CCC AAG TAT ATC TGT CAA AAT CAA CAT TCG ATC TCC ACT AAA CTC AAG CAA TCC TGT CAA AAA CCT CTG TTC GAA AAA TGT CAC TGC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320  
CCC CAA CTC CAA AAT CAT CAG ATG CCT CCT CAC TTG CCT TCA TTA CCT CAT TTT CTT GAA ACT AAG CAT CTT TCC AAA AAC TAT CCT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
CAG CCA AAG CAT CTC TTC TGC CCC ATG TTT TTC TAT CAA TAT CCA ACA ACG CAT CCT CAT TAC TCT GTC CTC CTC ACA CTT CCC (1160)

351 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his alu cys tyr ala lys val phe asp glu phe lys pro leu 380  
AAG ACA TAT CAA ACC ACT CTA CAG AAG TGC TGT TGT CCC CCT GCA CAT CCT CAT CAA TCC TAT CCC AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu leu phe glu tyr lys phe gln asn ala leu leu val arg 410  
CTG CAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT CCA CAG TAC AAA TTC CAG AAT CCC CTC TTA GTT CCT (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gly lys val gly ser lys cys cys lys his 440  
TAC ACC AAG AAA GTA CCC CAA CTG TCA ACT CCA ACT CTT GTA CAG CTC TCA ACA AAC CTA CCA AAA GTG CCC ACC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser 470  
CCT GAA CCA AAA ACA ATG CCC TGT CCA GAA CAG TAT CTA TCC GTG CTC AAC CAG TTA TGT GTG TTC CAT CAG AAA ACC CCA GTA AGT (1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 500  
CAC ACA CTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC ACG CCA CCA TCC TTT TCA CCT CTG CAA CTC CAT CAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val 530  
CAG TTT AAT CCT GAA ACA TTC ACC TTC CAT CCA GAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA ATC AAC AAA CAA ACT GCA CTT GTT (1700)

5  
10  
15  
20  
25  
30  
35

531 glu leu val lys his lys pro lys ala thr lys ala thr lys glu gln leu lys ala val cys asp phe ala ala phe val glu lys oys lys  
GAG CTC GTG AAA CAC AAG CCC AAG GCA ACA ACA GAG CAA CTC AAA CCT GTT ATG CAT CAT TTC CCT CCT TTT GTA GAG AAG TCC TCC AAG (1790)

558 559 560

540 567 570 580

561 ala asp asp lys glu thr cys phe ala glu glu gln lys leu val ala ala ser gln ala ala leu gln leu ter  
CCT CAC CAT AAG GAG ACC TCC TTT GCC GAG CAG GGT AAA AAA CTT GTT CCT GCA AGT CAA CCT GCC TTA TAA CATCACATTAAAG (1883)

ter ter

CATCTCAGCCTACCATCAGAAATAACAGAAAGAAATGAAGATCAAAACCTTATTCATCTGTCTTTCTTTTCGTTGGTGTAAAGCCCAACACCCCTGTCTAAAAACATATAATTTCTTTAA (2402)

TCATTTGCCCTCTTTCTCTGCTTCAATTATAAAAAATCGAAGAACTAA..... 20 .....AA (2078)

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

Met lys trp val tlu phe lle ser leu leu phe leu phe ser  
ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT ACC (30)

GCCTTTCTCTCTCTGTCACACCCACAGCCCTTTGCACACA

ser ala tyr ser arg gly val phe arg arg  
TCC CCT TAT TCC ACC GGT GTC TTT CGT CGA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

[illegible]

231 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu leu glu cys ala asp asp asp ala asp leu 260  
GTT TCC AAC TTA GTG ACA GAT CTT ACC AAA GTC CAC ACC GAA TCC TCC CAT CCA CAT CTG CTT CAA TGT CCT CAT CAC ACC GCG CAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile 289 290  
CCC AAG TAT ATC TGT GAA AAT CAA GAT TCC ATC TCC AGT AAA CTG AAG GAA TCC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TCC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320  
CCC GAA GTG CAA AAT GAT CAG ATG CCT CCT CAC TTG CCT TCA TTA CCT CCT CAT TTT GTT CAA AGT AAC GAT GTT TCC AAA AAC TAT CTT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
GAG GCA AAG CAT GTC TTC TTG CCC ATC TTT TTC TAT GAA TAT GCA ACA AGC CAT CCT GAT TAC TCT GTC CTC CTC CAC ACA CTT CCC (1160)

351 lys thr tyr glu thr thr leu glu lys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 380  
AAG ACA TAT CAA ACC ACT CTA GAG AAG TCC TGT GCT CCT GCA CAT CCT CAT GAA TCC TAT CCC AAA GTG TTC GAT CAA TTT AAA CCT CTT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu leu qly glu tyr lys phe gln asn ala leu leu val arg 410  
GTC CAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAT CTT TTT GAG CAC CTT GCA GAG TAC AAA TTC CAG AAT CCC CTC TTA GTT CTT (1360)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440  
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA CCA AAA GTG CCC ACC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qly lys thr pro val ser 470  
CCT CAA CCA AAA ACA ATG CCC TGT CCA GAA CAC TAT CTA TCC CTC CTC CAC AAC CAG TTA TGT GTG TTC CAT CAG AAA ACC CCA GTA AGT (1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 500  
CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA CCT CTG CAA GTC GAT CAA ACA TAC CTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val 530  
GAG TTT AAT CCT CAA ACA TTC ACC TTC CAT CCA CAT ATA TCC ACA CTT TCT GAG AAG GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

ter ter

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
Met lyo trp val leu phe lle ser leu leu phe leu phe ser
ATG AAG TGG GTA ACC TTT ATT TCC CTT TTT CTC TTT ACC (30)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
Ser ala tyr ser arg gly val phe arg arg asp ala his lyo ser glu val ala his arg phe lyo asp leu gly ala glu asn phe lys
TCG CCT TAT TCC ACC GGT GTG TTT CCG TTT CCG CCA GAT CCA CAC AAG AGT CAG GTT CCT CAT CCG TTT AAA GAT TTC GCA GAA GAA AAT TTC AAA (170)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
ala leu val leu lle ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr ala phe ala
CCC TTC GTG TTG ATT GCC TTT CCT CAG TAT CTT CAG CAG TGT CCA TTT CAA GAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT CTT CCT CAT GAG TCA CCT GAA AAT TGT CAG AAA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TCC ACA GTT CCA ACT CTT (350)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn ala cys phe leu gln his lys asp asp asn pro
CGT GAA ACC TAT CGT GAA ATG CCT CAC TCC CAC TCC TGT CCA AAA CAA CAA CCT CGG ACA AAT GAA TCC TTC TGC CAA CAC AAA CAT CAC AAC CCA (440)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu glu thr phe leu lys lys tyr leu try
AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT CAT GTG ATG TCC ACT CCT TTT CAT CAC AAT GAA CAG ACA TTT TTC AAA AAA TAC TTA TAT (530)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
glu lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr ala cys cys gln
CAA ATT GCC ACA ACA CAT CCT TAC TTT TAT CCC CCC GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT TTT ACA CAA TGT TCC CAA (620)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg ala leu lys cys
CCT GCT GAT AAA GCT GCC TCC CTG TTG CCA AAG CTC GAT GAA CTT CGG GAT CAA CGG AAG CCT TCG TCT CCC AAA CAG ACA CTC AAG TGT (710)
-10      0      10      20      30      40      50      60      70      80      90      100      110      120      130      140      150      160      170      180      190      200      210      220      230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala ala phe ala glu
CCC AGT CTC CAA AAA TTT CCA GAA ACA GCT TTC AAA CCA TCG GCA GTA GCT CGC ACC CAG ACA TTT CCC AAA GCT GAG TTT GCA CAA (800)

```



231 val ser lys leu val thr asp leu thr lys val hio thr glu oyo hio gly asp leu leu glu oys ala asp asp arg ala asp leu 260  
 GTT TCC AAG TTA GTG ACA CAT CTT ACC AAA GTC CAC ACC GAA TCC TCC CAT GCA CAT CTG CTT GAA TGT GCT GAT CAC AGC GCG CAC CTT (1890)  
  
 261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu oys qiu lys pro leu leu qiu lys ser his cys ile 289 290  
 CCC AAG TAT ATC TGT GAA AAT CAA CAT TCC ATC TCC ACC AAA CTG AAG GAA TCC TGT GAA AAA CCT CTC TTC GAA AAA TCC CAC TCC ATT (1980)  
  
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320  
 GCC GAA GTG GAA AAT CAT CAG ATG CCT CCT CAC TTG CCT TCA TTA CCT GCT GAT TTT GTT GAA AGT AAG CAT GTT TGC AAA AAC TAT CCT (1070)  
  
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
 GAG GCA AAG GAT GTC TTC TTC GGC ATG TTT TTG TAT GAA TAT GCA ACA AGC CAT CCT GAT TAC TCT GTC CTC CTC CTC ACA CTT GCC (1160)  
  
 351 lys thr tyr glu thr thr leu glu lys oys ala ala asp pro his qiu cys tyr ala lys val phe asp qiu phe lys pro leu 380  
 AAG ACA TAT GAA ACC ACT CTA CAG AAG TCC TGT GGT GCT GCA CAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)  
  
 381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe qiu gln leu qly glu tyr lys phe gln asn ala leu leu val arg 410  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTC TTA CTT CCT (1340)  
  
 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA CGA AAA GTG GCG ACC GAA TGT TGT AAA CAT (1430)  
  
 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu oys val leu his qiu lys thr pro val ser 470  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA CAA CAC TAT CTA TCC GTC GTC CTC AAC CAG TTA TGT GTC TTC CAT CAG AAA ACC CCA GTA AGT (1520)  
  
 471 asp arg val thr lys oys cys thr glu ser leu val asn arg arg pro oys phe ser ala leu glu val asp qiu thr tyr val pro lys 500  
 CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTC AAC AGC CCA CCA TCC TTT TCA CCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)  
  
 501 glu phe asn ala glu thr phe thr phe his ala asp ile oys thr leu ser glu lys glu arg qin ile lys lys qin thr ala leu val 530  
 CAG TTT AAT CCT GAA ACA TTC ACC TTC CAT CCA CAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA ATC AAC AAA CAA ACT CCA CTT GTT (1700)

201 201

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

# Restriction Endonuclease Map of Human Serum Albumin cDNA Clones

PHA36

